Industrial Property Prediction using Towhee and LAMMPS

Marcus G. Martin*, Aidan P. Thompson

Computational Materials and Molecular Biology,

Sandia National Laboratories, PO Box 5800, Mail Stop 0316, Albuquerque, NM 87185-0316

Abstract

Results are presented for Part 2 (density) and Part 3 (viscosity) of the First Industrial Fluid Properties Simulation Challenge. In both cases, the physical properties were calculated using existing published force fields not specifically tuned to the problem at hand. No assessment of the accuracy of our predictions was made until the experimental values for each problem set were announced at the end of the competition.

Liquid densities were computed for the Part 2 problem set using the Towhee Monte Carlo molecular simulation program and the Amber96, Charmm22, Compass, and OPLS-aa force fields. No single force field was able to provide parameters for all the molecules in the problem set, but the Amber96 force field had the best results of the four tested and a reasonable coverage of the problem set.

Viscosities were computed for the Part 3 problem set using the LAMMPS molecular dynamics code. The Towhee program was used to generate equilibrium starting configurations. Only one force field, OPLS-aa, was used. The predicted viscosities showed average deviation of about 35% from the experimental values. In cases where the experimental density is known, substantially better accuracy can be expected.

---

* Corresponding author: marmart@sandia.gov

Introduction

We were motivated to enter the First Industrial Fluid Properties Simulation Challenge (FIFPSC) for two reasons. First, we wanted to provide a baseline comparison of the results an industrial scientist could expect when using existing force fields not tuned specifically to the problem at hand. Second, we wanted to demonstrate the power and utility of the Towhee and LAMMPS codes which we provide to the public here at Sandia National Laboratories.

We initially planned on entering all three of the Challenge categories, but attempts to compute the phase envelope for Part 1 were hindered by extremely low acceptance rates for molecule transfer moves for nitroethane and propylene glycol monomethyl ether at the required temperatures.

Liquid Density Prediction

The liquid densities for Part 2 of the FIFPSC were computed using the MCCCS Towhee simulation package. [1] All simulations were performed in the isobaric-isothermal ensemble with the temperature and pressures specified in the contest information. Simulations were equilibrated for at least 20,000 Monte Carlo cycles (one cycle is N moves where N is the number of molecules in the system), and results are reported for simulations of 10,000 cycles. Standard deviations were computed by breaking the simulations into 5 blocks.

The Monte Carlo moves consisted of volume changes, coupled-decoupled configurational bias regrowths, translation of the center-of-mass, and rotation about the center-of-mass. In addition, single atom translation moves were performed for cyclohexane and pyridine as the configurational bias acceptance rate was low for cyclic molecules.

The coupled-decoupled configurational-bias algorithm used here is based on previous

2

work [2] with a few modifications. Flexible bond lengths were introduced as an additional decoupled selection (with 1000 trial choices for bond length selection) where the biasing energy was the sum of the vibration terms plus any bond-bond cross terms. Improper torsions and angle-angle terms were added into the energies computed in the bond angle selection. An additional biasing function was used during regrowths of cyclohexane in order to encourage the growth to form the ring.

The Amber96 [3], Charmm22 [4], Compass [5], and OPLS-aa [6] force fields were utilized in this study. These force fields are among the most commonly used in the simulation community and exist in a state of ongoing competition with each other. Amber96 and Charmm22 were parameterized for biomolecular simulations and therefore have particular emphasis on the amino acids. The biological origin of hydrocarbon feed-stocks means that a good number of industrially relevant functional groups are covered by these force fields. OPLS-aa also has a biological emphasis, but is more broadly focused on a variety of organic compounds, and therefore has a more complete coverage of the FIFPSC test suite. The final force field studied was the Compass force field which was designed for broad use in the chemical industry. Unfortunately, not all of the parameters for this force field are publicly available, as it is sold as a portion of the proprietary codes of Accelrys.

A 10 Å cutoff with analytical tail corrections was used for all four force fields, and coulombic interactions were computed using the Ewald sum method. Charge assignments were either made based upon the formal rules of the force field (Compass), or inferred from comparison to similar molecules that have been published in the literature (Amber96, Charmm22, OPLS-aa). A complete list of the charge assignments and atom types is provided in the supplementary information.

The computed liquid densities are shown in Table 1. In general there is good agreement between the various force fields for those molecules where comparison is possible. The

3

experimental values were presented in the introduction to this issue.

Using this data we can assess how well each force field performed according to a number of criteria. First, we consider the generality of the force fields based upon the number of molecules for which parameters were available. No single force field has published parameters suitable for all nine of the molecules in the FIFPSC data set. Ordering the force fields by coverage yields OPLS-aa (8), Amber96 (7), Charmm22 (4), and then Compass (3).

We also consider the accuracy of the different force fields by computing the root mean square (RMS) deviation between predicted and experimental densities as shown in Table 2. Looking at the overall RMS deviation for each force field shows that Amber96 and Charmm22 are better than Compass and OPLS-aa. However, this is not a fair comparison as the different force fields did not all have an answer for each of the entries. To eliminate the possibility that some of the entries were more difficult than others we also look only at the simulations where all of the force fields had parameters (cyclohexane and isopropanol). Here we can see that Amber96 is better than the other three (which all do about equally well).

The results for water deserve special mention. The newest, and most expensive, water model tested (OPLS-aa TIP5P) [7] does not give improved agreement with experiment compared to its own predecessors (Amber96 and Charmm22 use variants of the TIP3P model). Also surprising are the differences seen between the Amber96 and Charmm22 versions of the TIP3P model. The Amber96 version is true to the original TIP3P in that there are no Lennard-Jones terms on the hydrogens, while Charmm22 does have some Lennard-Jones terms on the hydrogens. The added terms make the Charmm22 water model more attractive, and therefore result in higher densities.

Viscosity Prediction

The liquid viscosities were obtained from molecular dynamics simulations run using the LAMMPS classical molecular dynamics code. [8,9,10] There exist a variety of methods for calculating the shear viscosity of liquids from molecular dynamics simulation. In the limit of infinite computer resources, all of these methods are formally equivalent. In comparative studies, some of the more popular methods have been shown to yield statistically equivalent estimates of viscosities for specific systems. [11,12] The methods can be divided into two main classes: equilibrium (EMD) and non-equilibrium (NEMD). Equilibrium methods sample the time-dependence of thermal fluctuations in the stress or momentum flux tensors, which can be related to the viscosity using linear response theory. [13] Non-equilibrium methods sample the response of the system to an applied external field (shear rate or shear stress). Both classes have their strengths and weaknesses. EMD methods are easier to implement in a general-purpose MD code. A single EMD simulation can be used to measure other properties besides the viscosity. In EMD methods, there are no adjustable parameters controlling the accuracy or precision of the viscosity measurement. NEMD methods require substantial code modification and can not be used to simultaneously measure other properties. Usually, several runs must be made in order to identify the best choice of external field strength. Once the optimal choice of field strength is known, NEMD methods tend to outperform EMD methods in terms of the CPU time required to achieve a specified level of precision in the viscosity. However, it is important to note that the single biggest influence on the CPU requirement is the viscosity of fluid itself. This is because viscosity is controlled by the longest relaxation time in the fluid, which is usually associated with molecular rotations. In the case of EMD, in order to obtain good statistical precision, the simulation must be long enough to sample many such rotations. In the case of NEMD, in order to avoid systematic deviations from the zero-frequency viscosity, the shear rate must be much less than the rotational rate constant. In both cases, larger viscosities require longer simulations.

After weighing all of these factors, we selected an EMD method based on fluctuations in the stress tensor. [12,13] This has been found to work well for fairly large polyatomic molecules such as n-hexadecane [14] and HMX. [15] The EMD simulations were performed in the NVT ensemble using the LAMMPS molecular dynamics code. [8,9,10] Equilibrium starting configurations were obtained by using the Towhee code to perform an isobaric-isothermal Monte Carlo simulation. As well as providing the starting configuration, Towhee also automatically generated all of the force field data required by LAMMPS. This eliminated the time-consuming and error prone task of constructing force field data by hand. Because of the long runs required to obtain an accurate estimate of viscosity from the MD simulations, we decided to test just one force field, and we chose OPLS-aa.

Each of the four mixtures was represented by roughly 1000 atoms. The cubic box dimensions were approximately 22 Å. The actual densities used for each system are given in Table 3. We have included errors calculated by comparison with the experimental densities announced at the end of the competition. The simulation densities were slightly too high for n-nonane, isopropanol and the 25%/75% mixture. In contrast, the density used for the 50%/50% mixture was lower than experiment. These deviations are comparable to the OPLS-aa RMS errors for density prediction observed in the previous section.

Having obtained a starting configuration, NVT EMD simulations were carried out for each system using four nodes (Dec alpha 500 MHz) of the CPlant high performance computing platform at Sandia National Laboratories. A multiple-timestep velocity-Verlet integration scheme was used with 1 fs timestep for bonded interactions and 2 fs timestep for non-bonded and long-range Coulombic interactions. Coulombic interactions were calculated using particle-mesh Ewald summation on an 8x8x8 grid with a 10 Å cut-off for the real-space contribution. The decay constant for the Gaussian charges was set to 0.237 Å^{-1}. The simulation throughput was about 40 fs/second.

Each run was equilibrated for 2 ps and followed by a production run of at least 20 ns. To obtain the macroscopic viscosity, we calculated the components of the symmetric traceless stress tensor at each timestep following Mondello and Grest [14] and Daivis and Evans [12]. The time-integral of each shear stress component (L_{xx}, L_{yy}, L_{zz}, L_{xy}, L_{xz}, L_{yz}) was written to disk every 200 fs. Each of these quantities is one component of a six-dimensional random walk. For each component we calculated the average mean-square-displacement $< L_{\alpha\beta}(\Delta t) >$ for a series of successively larger time displacements Δt_i. The values of Δt_i used were powers of two multiples of 200 fs. For each Δt_i, the average was taken over all consecutive non-overlapping time intervals. For each component, the time-dependent shear viscosity $\eta_{\alpha\beta}(\Delta t)$ was then expressed as a numerical derivative of $< L_{\alpha\beta}(\Delta t) >$.

$$\eta_{\alpha\beta}(1/2(\Delta t_i + \Delta t_{i+1})) = \frac{V}{2kT} q_{\alpha\beta} \frac{< L_{\alpha\beta}(\Delta t_{i+1}) > - < L_{\alpha\beta}(\Delta t_i) >}{\Delta t_{i+1} - \Delta t_i} \qquad [1]$$

where V is the simulation box volume, T is the temperature, k is Boltzmann's constant. $q_{\alpha\beta} = 1$ for $\alpha \neq \beta$ and $q_{\alpha\beta} = 3/4$ for $\alpha = \beta$. All six $\eta_{\alpha\beta}(\Delta t)$ are equivalent and in the limit of large Δt, they should approach a constant value, the zero frequency shear viscosity.

Figure 1 shows the time-dependent shear viscosity averaged over all six components equally. Error bars were obtained by dividing each simulation into ten blocks and calculating the variance over blocks and components, 60 data points in total. Subsequent analysis (not shown) indicated that the variance of the viscosity obtained from L_{xx}, L_{yy} and L_{zz} was one half greater than the variance of the viscosity obtained from L_{xy}, L_{xz} and L_{yz}. This is to be expected, as the former three quantities are not independent, due to the constraint that they sum to zero. Defining the overall average viscosity as an inverse variance-weighted average over the viscosity for each component yields the formula originally proposed without justification by Daivis and Evans [12]. Using their formula instead of a simple average over components did not change the results significantly.

In order to compute the macroscopic (zero frequency) shear viscosity, we needed to estimate the long-time limit of the time-dependent shear viscosity. This was done by first using the graphs in Fig. 1 to identify Δt_{lower}, such that for all points for which $\Delta t > \Delta t_{lower}$, the shear viscosity did not deviate significantly from the apparent long-time value. The identification of Δt_{lower} was unambiguous, except in the case of isopropanol, which had the highest viscosity and the longest relaxation time. The values of Δt_{lower} used are given in Table 3. All Δt points greater than Δt_{lower} were then used in the long time average. Since the statistical uncertainty of the points increases with Δt, we used a reciprocal variance weighted average over the long-time region. This heavily weighted the shorter time points, so that the decision to include or exclude the longest time points had a negligible effect on the averages.

The final viscosity predictions are shown in Table 4. The standard error estimate included with each value indicates the expected standard deviation of the the underlying distribution. Isopropanol had the largest standard error, both in absolute and relative terms. We expect that in all four cases, systematic errors due to the simulation method and data analysis method is less than the statistical uncertainty.

Relative errors were calculated by comparison with the experimental viscosities announced at the end of the competition. At first glance, the errors in the predicted viscosities appeared to be large and also highly variable. However, further analysis revealed a more consistent picture. We see that in the three cases where the density was overpredicted, the viscosity was also overpredicted, by an average of about 35%. In the one case where the density was underpredicted, the viscosity came out very close to the experimental value. Hence we expect that if the experimental densities had been available, all of the viscosities would have been overpredicted by roughly similar amounts, and the average relative error would be considerably improved.

Conclusions

We attempted to mimic the approach of an industrial scientist when confronted with the problems for this contest. We utilized publicly available software and force fields to make our predictions and found reasonable agreement with the experimental data. An industrial scientist following our approach could expect an error of around 0.04 g/ml when predicting liquid densities. In the case of viscosity prediction, the observed average error was 35%. However, if accurate experimental densities are available, the error in the viscosity predictions could be substantially reduced.

Acknowledgments

References

[1] http://www.cs.sandia.gov/projects/towhee/.

[2] M. G. Martin, J. I. Siepmann, J. Phys. Chem. B 103 (1999) 4508-4517.

[3] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, P. A. Kollman, J. Am. Chem. Soc. 117 (1995) 5179-5197.

[4] A. D. MacKerell Jr., D. Bashford, M. Bellott, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, M. Karplus, J. Phys. Chem. B 102 (1998) 3586-3616.

[5] H. Sun, J. Phys. Chem. B 102 (1998) 7338-7364.

[6] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, J. Am. Chem. Soc. 118 (1996) 11225-11236.

[7] M. W. Mahoney, W.L. Jorgensen, J. Chem. Phys. 112 (2000) 8910-8922.

[8] http://www.cs.sandia.gov/ sjplimp/lammps.html.

[9] S. J. Plimpton, R. Pollock, and M. Stevens, in Proc of the Eighth SIAM Conference on Parallel Processing for Scientific Computing, Minneapolis, MN, March (1997).

[10] S. J. Plimpton, J. Comp. Phys., 117 (1995), 1-19.

[11] B. Hess, J. Chem. Phys. 116 (2002) 209-217.

[12] P. J. Daivis, D. J. Evans, J. Chem. Phys. 100 (1994) 541-547.

[13] D. J. Evans and G. P. Morriss, Statistical Mechanics of Non-Equilibrium Liquids Academic Press 1990.

[14] M. Mondello, G. S. Grest, J. Chem. Phys. 106 (1997) 9327-9336.

[15] D. Bedrov and G. D. Smith, J. Chem. Phys. 112 (2000) 7203-7208.

Figure 1 Plot of time-dependent shear viscosity for pure n-nonane, pure isopropanol, and the 50/50 and 25/75 mixtures. The error bars for each shear viscosity value indicate the expected standard deviation of the underlying distribution.

Table 1 Liquid Densities computed using the Towhee Monte Carlo simulation code. Subscripts show the standard deviation of the last digits. Experimental data was provided by the contest organizers.

Molecule	Temp. [K]	Pressure [MPa]	Amber [g/ml]	Charmm [g/ml]	Compass [g/ml]	OPLS-aa [g/ml]	Average [g/ml]	Expt. [g/ml]
(a) water	293	0.1	0.996_7	1.051_9	N/A	0.987_8	1.011	0.9982
	423	2	0.857_{11}	0.917_4	N/A	0.820_9	0.865	0.9180
(b) cyclohexane	300	0.1	0.764_4	0.745_8	0.751_7	0.762_3	0.756	0.7721
	400	20	0.693_5	0.673_7	0.673_{13}	0.669_8	0.677	0.7029
(c) isopropanol	298.15	0.1	0.773_4	0.761_4	0.771_4	0.772_5	0.770	0.7819
	400	5	0.655_{11}	0.619_{14}	0.633_{10}	0.639_4	0.637	0.6804
(d) diethanol amine	330	0.1	1.052_3	N/A	N/A	1.009_3	1.03	1.0727
	400	5	1.022_4	N/A	N/A	0.986_4	1.004	1.0252
(e) 1,2,3-trichloropropane	290	0.1	N/A	N/A	N/A	1.385_{11}	1.385	1.3933
	400	1.5	N/A	N/A	N/A	1.215_{13}	1.215	1.2459
(f) triethylene glycol	280	2	1.077_3	N/A	1.048_1	1.022_4	1.049	1.1350
	310	0.1	1.088_2	N/A	1.029_3	0.991_2	1.036	1.1107
(g) pyridine	298	0.1	N/A	N/A	N/A	0.941_5	0.941	0.9782
	375	10	N/A	N/A	N/A	0.858_4	0.858	0.9090
(h) water and	298	0.1	1.040_4	N/A	N/A	N/A	1.040	1.0205
choline chloride	305	1	1.012_5	N/A	N/A	N/A	1.012	1.0068
(i) water and methanol	325	0.1	0.866_6	0.882_4	N/A	0.866_3	0.871	0.8977
	400	10	0.780_6	0.808_4	N/A	0.761_6	0.783	0.8288

Table 2 Root mean square errors in density for all systems studied, and for the subset of cyclohexane (b) and isopropanol (c). The subscripts show the average standard deviation in the last digit of the data points used to compute the RMS.

Force Field	Total RMS [g/ml]	(b) and (c) RMS [g/ml]
Amber96	0.030_5	0.015_6
Charmm22	0.034_7	0.038_8
Compass	0.055_6	0.030_9
OPLS-aa	0.059_6	0.028_5
Average	0.040	0.027

Table 3 Mass densities used in the NVT molecular dynamics simulations of the n-nonane/isopropanol mixtures. The errors were calculated by comparison with the experimental values.

System	Density	Error
	[g/ml]	[g/ml]
(a) n-nonane	0.718	0.005
(b) isopropanol	0.806	0.027
(c) 50/50	0.707	-0.022
(d) 25/75	0.762	0.016

Table 4 Viscosities for the OPLS-aa force field. Subscripts show the standard error in the last digits. The relative errors were calculated by comparison with the experimental values.

System	Δt_{lower}	Viscosity	Rel. Error
	[ns]	[mPa s]	[%]
(a) n-nonane	0.1	0.91_5	40
(b) isopropanol	0.1	3.0_3	51
(c) 50/50	0.1	0.78_6	3
(d) 25/75	0.05	1.32_5	27

Supplementary Table 1 Atom types and charge assignments

Atom	Amber-96		Charmm-22		Compass		OPLS-aa	
	Type	Charge	Type	Charge	Type	Charge	Type	Charge
water								
O	OW	-0.82	OT	-0.834	-	-	OWt5p	0.0
H	HW	0.41	HT	0.417	-	-	HW	0.241
lone pair	-	-	-	-	-	-	L5p	-0.241
cyclohexane								
C	CT	-0.2	CT	-0.18	c4	-0.106	CT	-0.12
H	HC	0.1	HC	0.09	h1	0.053	HC 1	0.06
isopropanol								
C_1	CT	-0.2438	CT3	-0.27	c4	-0.159	CT	-0.18
C_2	CT	0.364	CT1	0.14	c4o	0.107	CT	0.205
C_3	CT	-0.2438	CT3	-0.27	c4	-0.159	CT	-0.18
O (C_2)	OH	-0.6761	OH1	-0.66	o2h	-0.58	OHm	-0.683
H (O)	HO	0.4102	H	0.43	h1o	0.42	HO	0.418
H (C_1)	HC	0.0642	HA	0.09	h1	0.053	HC 1	0.06
H (C_2)	H1	0.0043	HA	0.09	h1	0.053	HC 1	0.06
H (C_3)	HC	0.0642	HA	0.09	h1	0.053	HC 1	0.06
diethanol amine								
H (O)	HO	0.4275	-	-	-	-	HO	0.418
O	OH	-0.6546	-	-	-	-	OHm	-0.683
C_1 (O)	CT	0.2117	-	-	-	-	CT	0.145
C_2 (C,N)	CT	-0.0249	-	-	-	-	CT	0.12
N	N	-0.4157	-	-	-	-	NT 2	-0.78
H (C_1)	H1	0.0352	-	-	-	-	HC 1	0.06
H (C_2)	H1	0.0209	-	-	-	-	HC 1	0.06
H (N)	H	0.2719	-	-	-	-	H	0.30
1,2,3-trichloropropane								
C_1	-	-	-	-	-	-	CT	0.0
C_2	-	-	-	-	-	-	CT	0.06
C_3	-	-	-	-	-	-	CT	0.0
Cl	-	-	-	-	-	-	Cl	-0.12
H	-	-	-	-	-	-	HC 1	0.06
triethylene glycol								
H (O)	HO	0.40	-	-	h1o	0.42	HO	0.44
O (H,C)	OH	-0.65	-	-	o2h	-0.58	OHp	-0.7
$C_{1,6}$	CT	0.18	-	-	c4o	0.054	CT	0.14
$C_{2,5}$	CT	0.18	-	-	c4o	0.054	CT	0.14
O (C,C)	OS	-0.35	-	-	o2e	-0.32	OS	-0.4
$C_{3,4}$	CT	0.18	-	-	c4o	0.054	CT	0.14
H ($C_{1,6}$)	H1	0.01	-	-	h1	0.053	HC 1	0.06
H ($C_{2,5}$)	H1	0.01	-	-	h1	0.053	HC 1	0.03
H ($C_{3,4}$)	H1	0.01	-	-	h1	0.053	HC 1	0.06
pyridine								
N_1	-	-	-	-	-	-	NC	-0.678
$C_{2,6}$	-	-	-	-	-	-	CA	0.473
$C_{3,5}$	-	-	-	-	-	-	CA	-0.447
C_4	-	-	-	-	-	-	CA	0.227
H ($C_{2,6}$)	-	-	-	-	-	-	HA	0.012
H ($C_{3,5}$)	-	-	-	-	-	-	HA	0.155
H (C_4)	-	-	-	-	-	-	HA	0.065

Supplementary Table 2 Atom types and charge assignments

Atom	Amber-96		Charmm-22		Compass		OPLS-aa	
	Type	Charge	Type	Charge	Type	Charge	Type	Charge
choline								
H (O)	HO	0.4275	-	-	-	-	-	-
O (H,C)	OH	-0.6546	-	-	-	-	-	-
C_1	CT	0.2117	-	-	-	-	-	-
C_2	CT	-0.0249	-	-	-	-	-	-
N	N3	-0.3854	-	-	-	-	-	-
$C_{3,4,5}$	CT	-0.0249	-	-	-	-	-	-
H (C_1)	H1	0.0352	-	-	-	-	-	-
H (C_2)	HP	0.13	-	-	-	-	-	-
H ($C_{3,4,5}$)	HP	0.13	-	-	-	-	-	-
methanol								
C	CT	0.1215	CT3	-0.04	-	-	CT	0.145
O	OH	-0.6546	OH1	-0.66	-	-	OHm	-0.683
H (O)	HO	0.4275	H	0.43	-	-	HO	0.418
H (C)	H1	0.0352	HA	0.09	-	-	HC 1	0.04